IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Ghislain Madiot>
<February 2nd 2026>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- **Objectives**

  This project applies data science and machine learning to analyze SpaceX Falcon 9 launch data and identify the key factors influencing first-stage landing success, a critical driver of launch cost reduction.

- **Methodologies**

  Historical launch data was collected from the SpaceX API and Wikipedia, cleaned and engineered, and analyzed using EDA, SQL, interactive maps, and dashboards to uncover patterns related to launch site, payload mass, orbit type, and flight history.

- **Results**

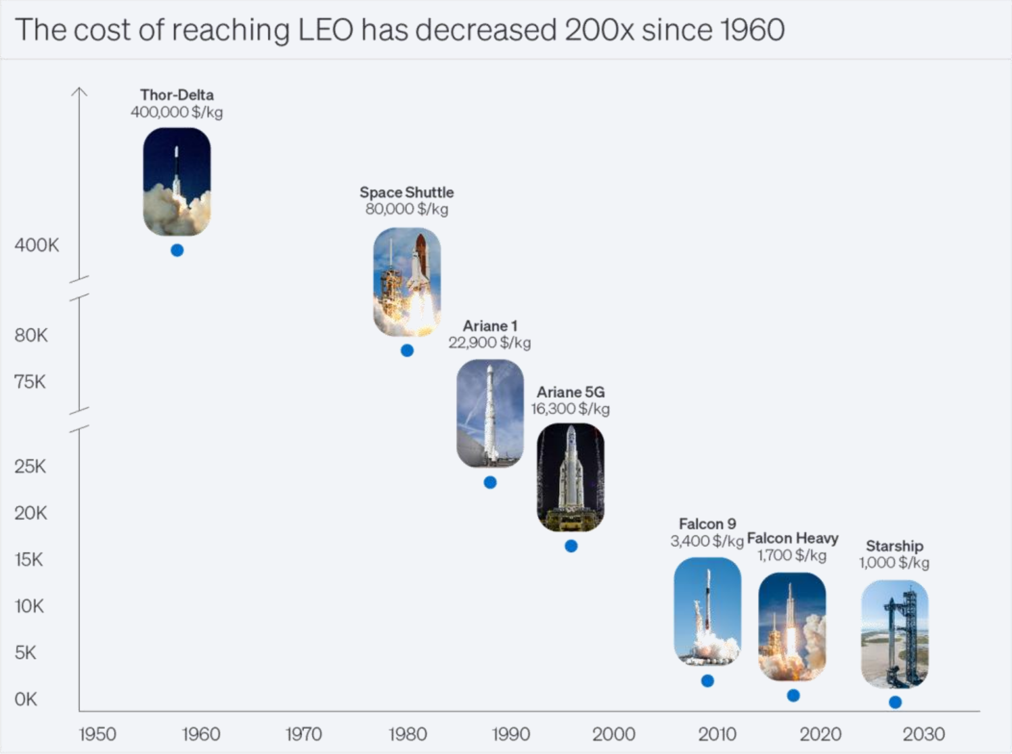
  Results show that landing success improves significantly over time, varies by orbit type, and is highest at KSC LC-39A; heavier payloads exhibit greater variability but maintain high success in later missions.

  Several classification models were evaluated, with the Decision Tree classifier achieving the best performance (~94% accuracy), demonstrating strong predictive capability for landing success.

Introduction

- Background and context
  - The cost of reaching Low Earth Orbit (LEO) has dropped by nearly 200x since the 1960s, driven largely by reusable rocket technology.
  - SpaceX’s recovery of the first-stage booster has transformed launch economics, reducing costs to ~$62M per launch, compared to over $165M for non-reusable alternatives.
  - Because launch cost depends heavily on whether the first stage successfully lands, predicting landing outcomes is critical.

- Objectives
  - This project uses data science and machine learning to analyze historical launch data and identify the factors that determine first-stage landing success.



Evolution of Cost to reach LEO since 1960. Source: https://colossus.com/article/is-space-investable/#ref-8

Section 1

Methodology

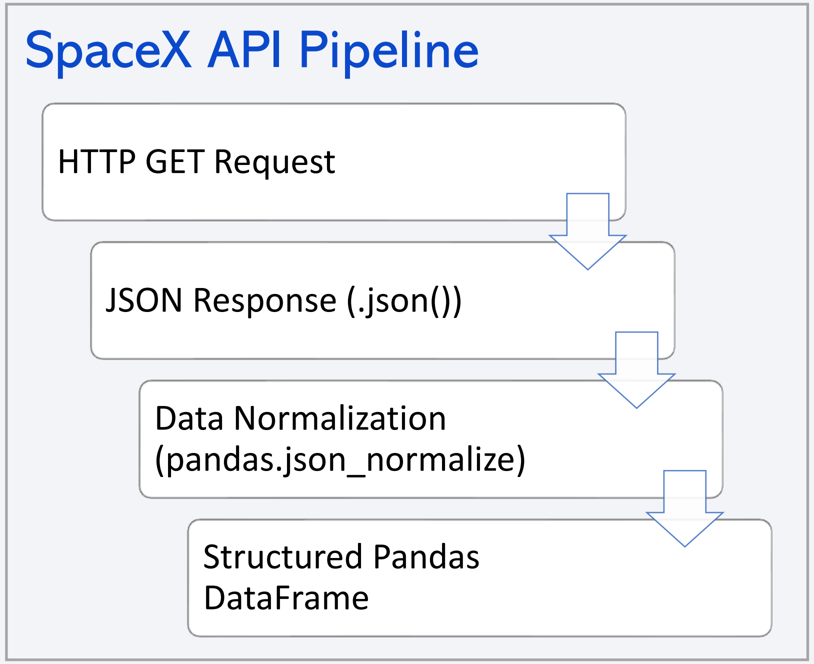
Methodology

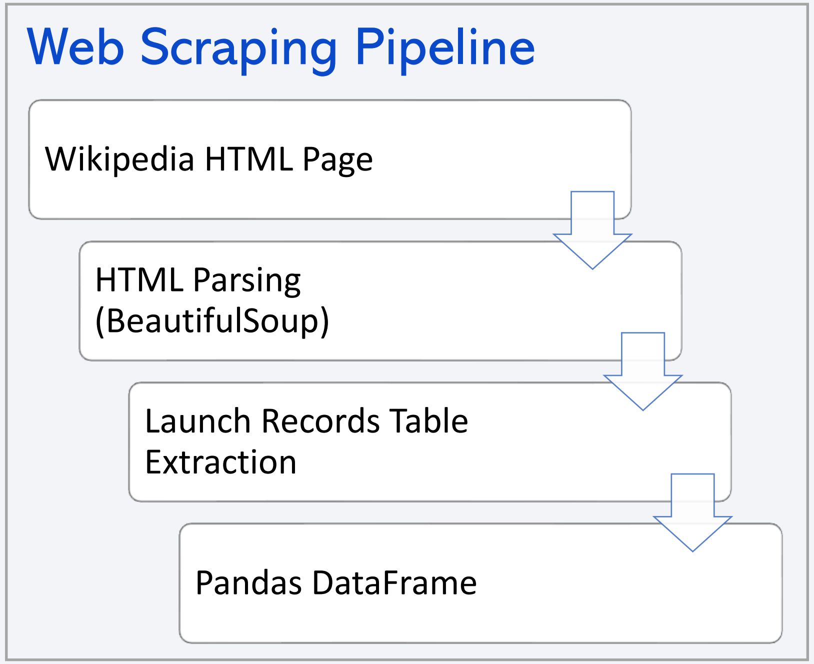
Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - One-hot encoding was applied to categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Key Data Sources
  - **SpaceX REST API** – launch, rocket, payload, and landing data
  - **Web scraping from Wikipedia** – historical Falcon 9 launch records
- Two Data Collection Processes





Data Collection – SpaceX API

- **Source**
  - **SpaceX REST API:** Publicly available, structured launch data
- **Methodology**
  - Data retrieved using **HTTP GET requests**
  - API responses returned in **JSON format**
  - JSON objects converted into structured tables using:
    - response.json()
    - pandas.json_normalize()
- **Data Extracted**
  - Launch date and outcome
  - Rocket and payload information
  - Launch site and orbit type
  - First-stage landing results

Code Snippet

```python
# SpaceX API endpoint
url = "https://api.spacexdata.com/v4/launches/past"

# GET request
response = requests.get(url)
response.status_code
# Convert JSON response to DataFrame
data = response.json()
df_launches = pd.json_normalize(data)

# Preview data
df_launches.head()
```

GitHub URL : https://github.com/GMadiot/IBM-Data-Science-and-Machine-Learning-Capstone-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- **Source**
  - **Wikipedia: **Publicly available, Falcon 9 historical launch records
- **Methodology**
  - HTML page retrieved using HTTP requests
  - Page parsed with BeautifulSoup
  - Launch records extracted from HTML tables
  - Tables converted into Pandas DataFrames
- **Data Extracted**
  - Launch site
  - Rocket and payload information
  - Launch site and orbit type
  - Orbit

Code Snippet

```
[4]: static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

Next, request the HTML page from the above URL and get a `response` object

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
[5]: # use requests.get() method with the provided static_url
     # assign the response to a object
     response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```
[6]: # Use BeautifulSoup() to create a BeautifulSoup object from a response text content
     headers = {
         "User-Agent": "Mozilla/5.0 (compatible; DataScienceBot/1.0)"
     }

     response = requests.get(static_url, headers=headers)
     soup = BeautifulSoup(response.text, "html.parser")
```

Print the page title to verify if the `BeautifulSoup` object was created properly

```
[20]: # Use soup.title attribute
      print(soup.title.string)

      soup.title

List of Falcon 9 and Falcon Heavy launches - Wikipedia
```

GitHub URL : https://github.com/GMadiot/IBM-Data-Science-and-Machine-Learning-Capstone-Project/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

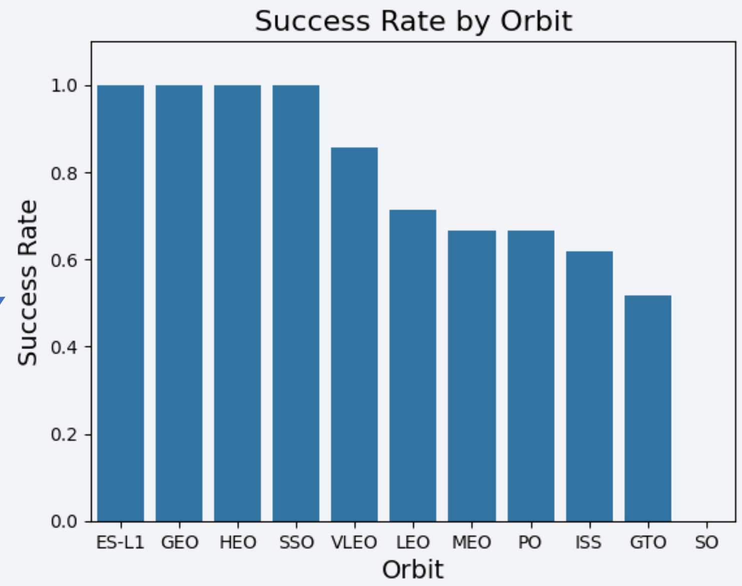
- **Key Processing Steps**
  - Selected relevant features related to launch and landing outcomes
  - Converted date and categorical variables to appropriate formats
  - Created binary class label for first-stage landing success
  - Handled missing and inconsistent values
  - Filtered and standardized datasets for consistency
- **Feature Engineering**
  - Landing outcome encoded as:
    - 1 → Successful landing
    - 0 → Failed landing

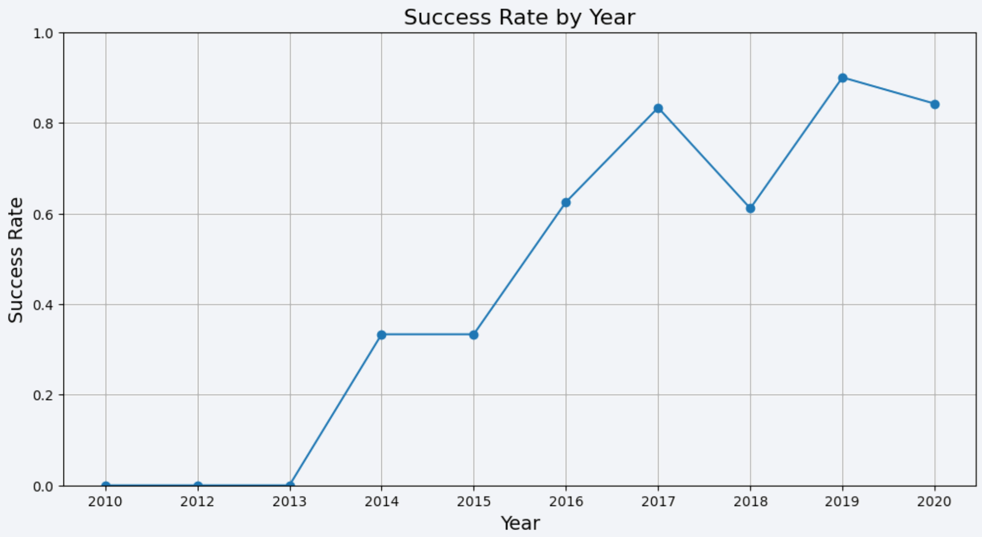
1. Raw API & Web Data
2. Column Selection
3. Data Type Conversion
4. Missing Value Handling
5. Feature Engineering
6. Clean Analysis Dataset

GitHub URL : https://github.com/GMadiot/IBM-Data-Science-and-Machine-Learning-Capstone-Project/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- We explored the data by visualizing the relationship between different variables:
  - flight number and launch Site,
  - payload and launch site,
  - **success rate of each orbit type,**
  - flight number and orbit type,
  - **the launch success yearly trend.**





GitHub URL : https://github.com/GMadiot/IBM-Data-Science-and-Machine-Learning-Capstone-Project/blob/main/edadataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

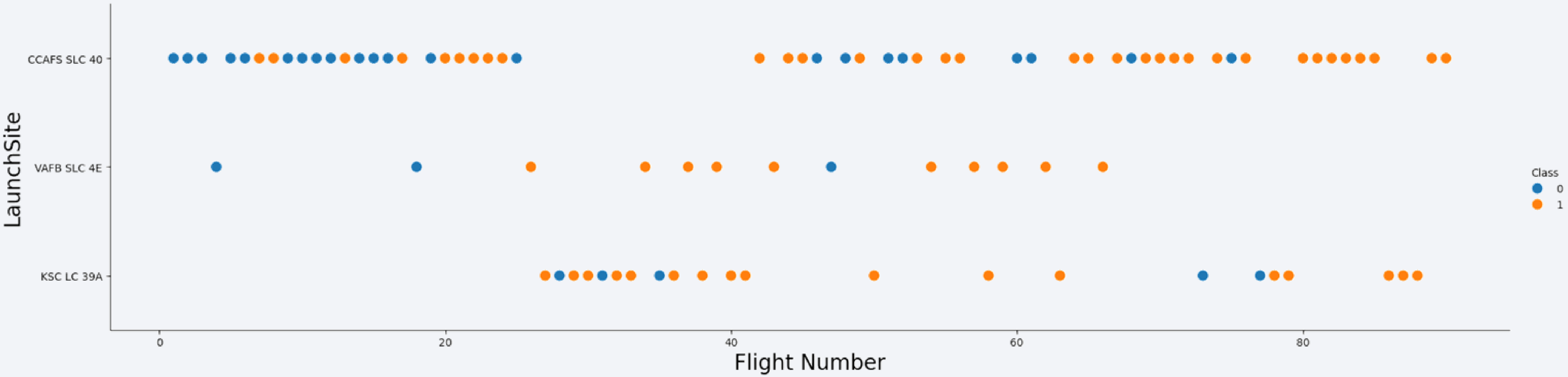
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

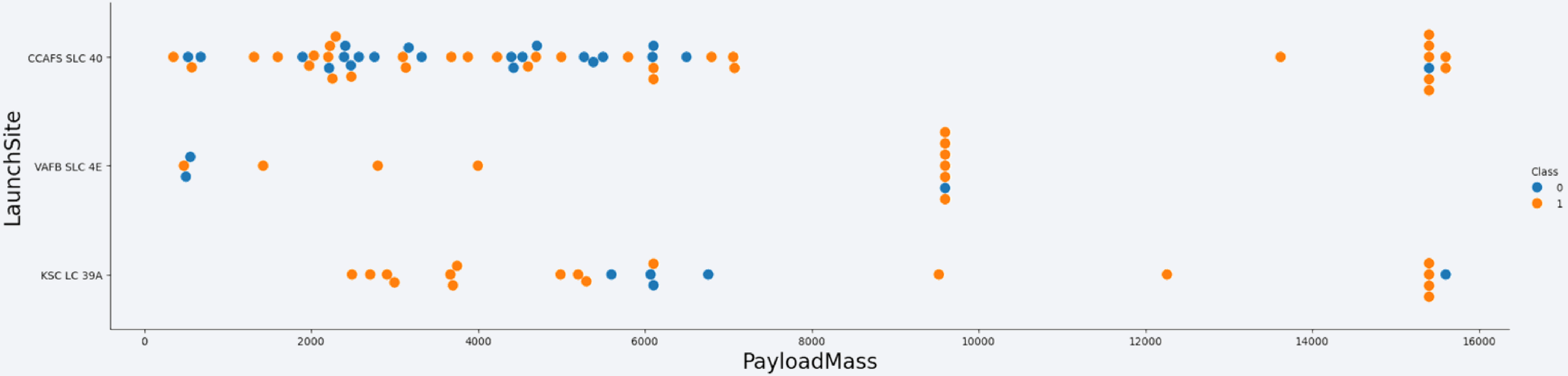
Insights drawn from EDA

Flight Number vs. Launch Site



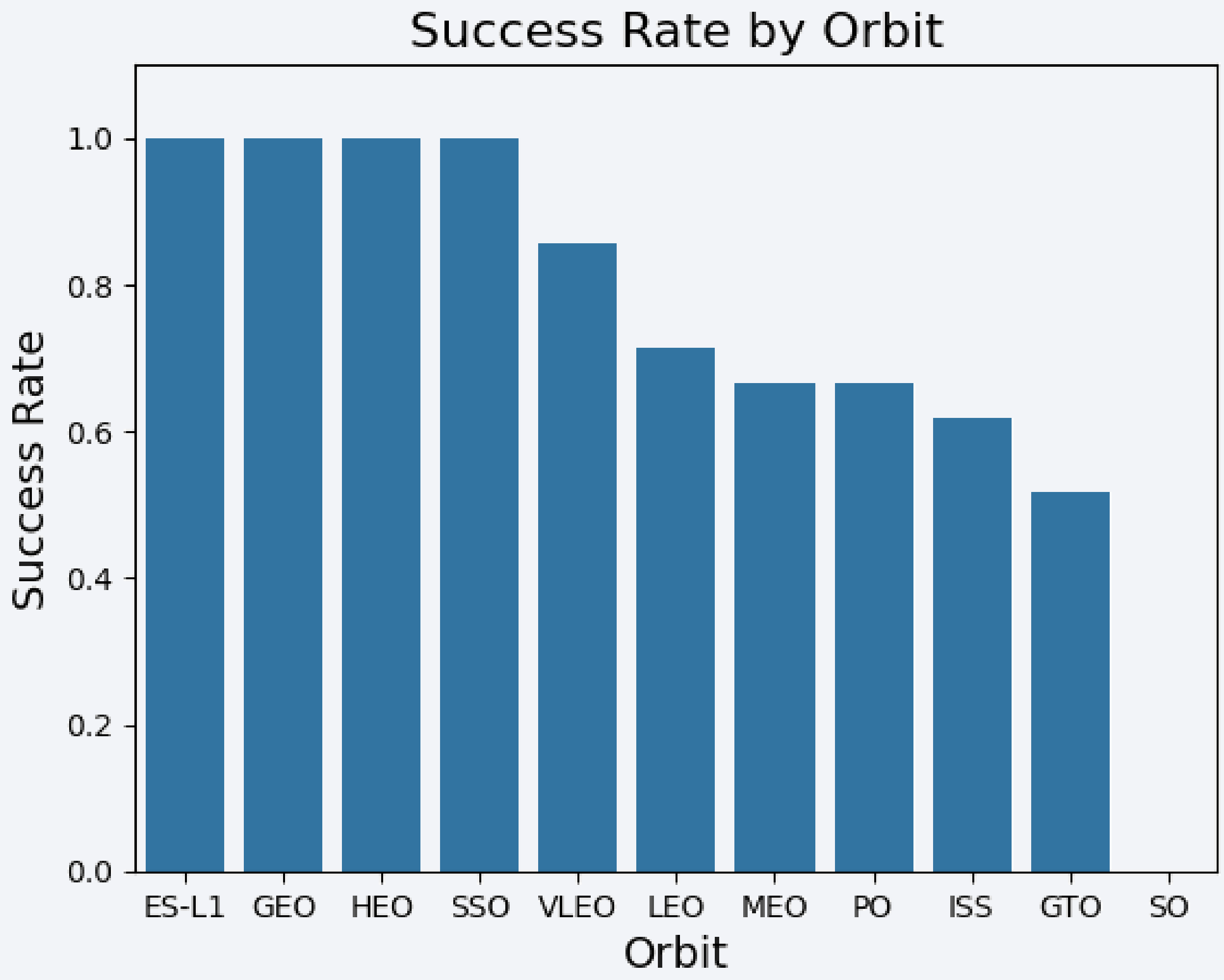
- Early flights show mixed success and higher failure rates
- Landing success increases with flight number across all launch sites
- CCAFS SLC-40 dominates launch frequency and success
- KSC LC-39A shows high success after operational maturity

Payload vs. Launch Site



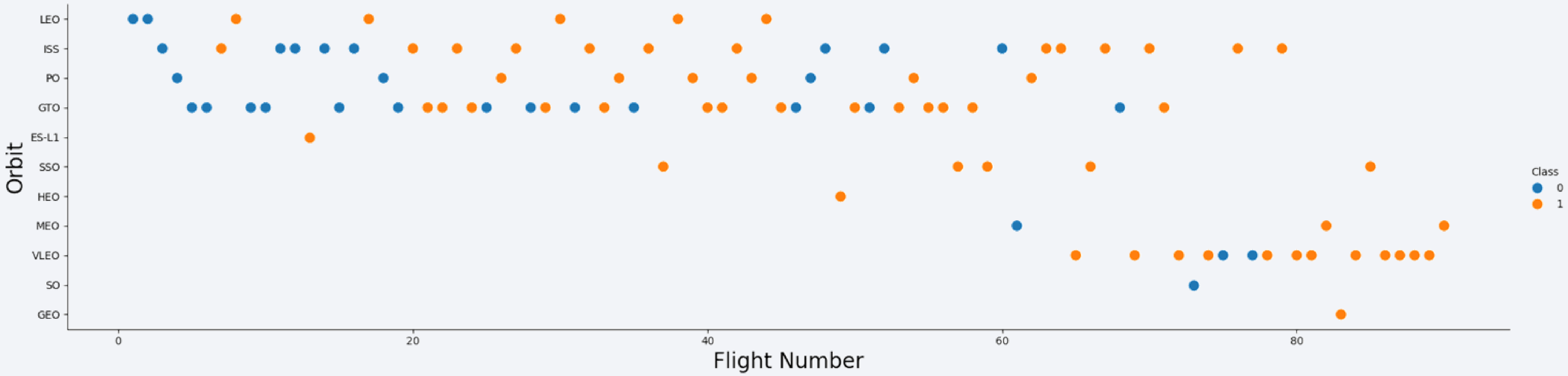
- Successful landings occur across a wide range of payload masses
- Heavier payloads show higher variability in landing outcomes
- CCAFS SLC-40 supports the widest payload mass range
- KSC LC-39A is associated with higher payload missions and high success rates

Success Rate vs. Orbit Type



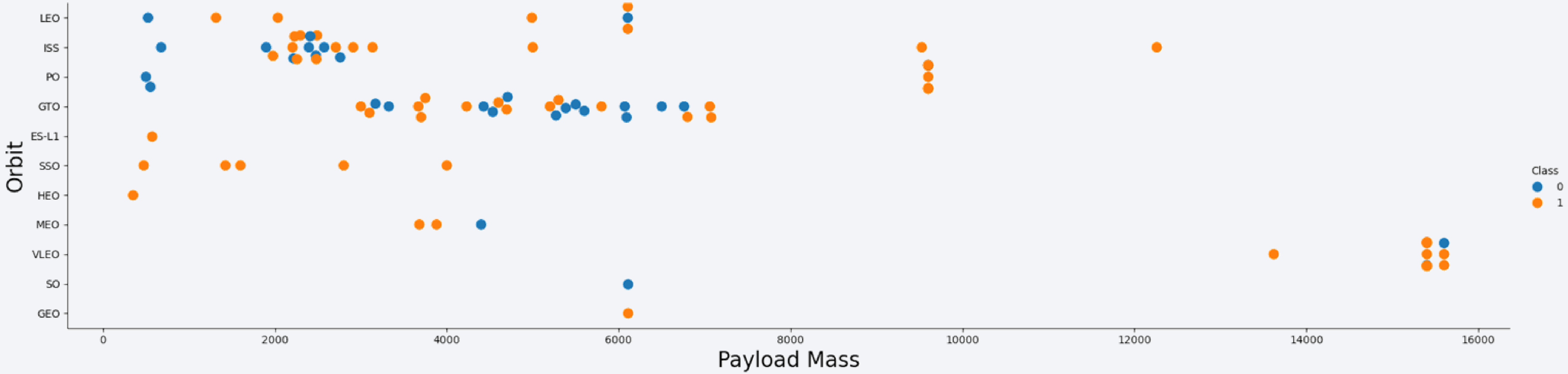
- Certain orbits achieve consistently high success rates
- LEO-related orbits show strong landing performance
- Success varies significantly by orbit

Flight Number vs. Orbit Type



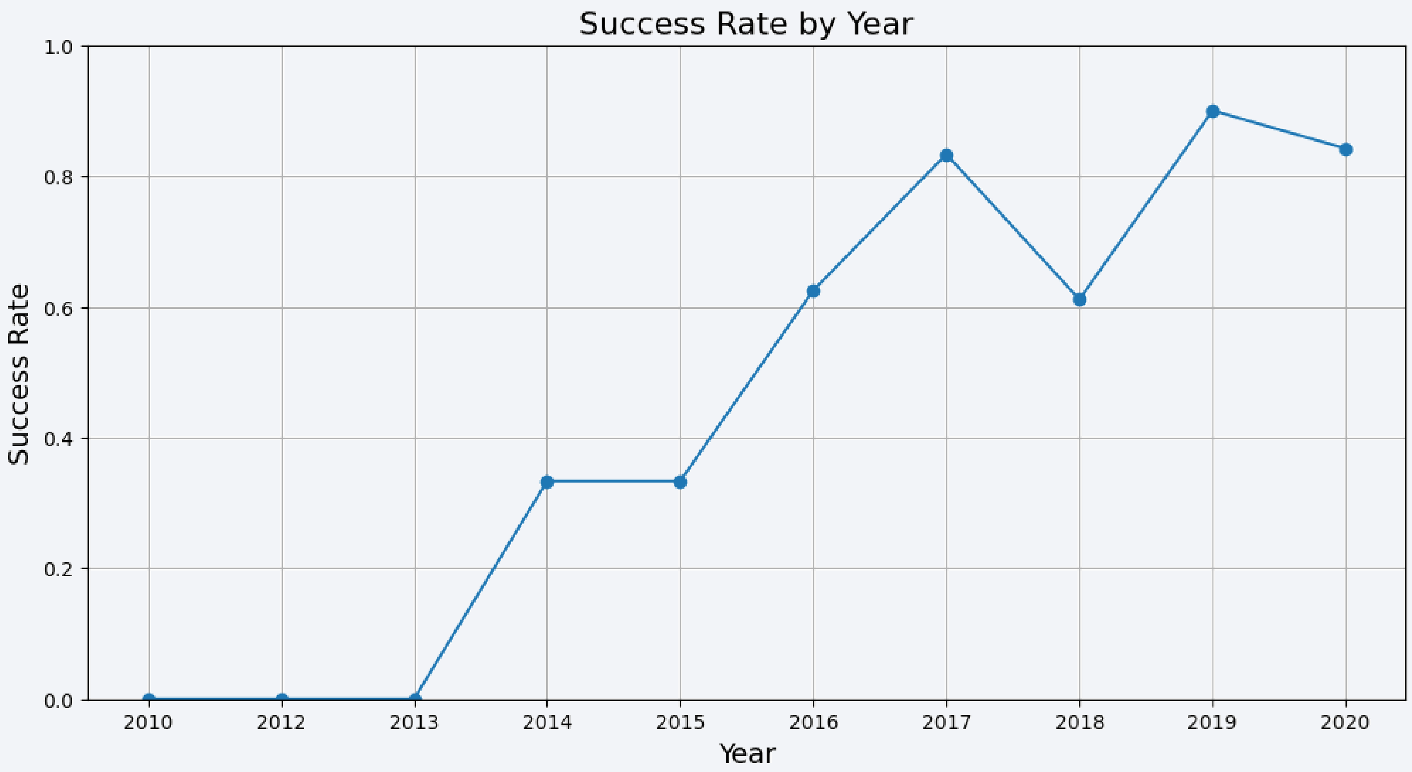
- Early flights show mixed landing outcomes across orbit types
- Certain orbits (e.g., LEO, ISS, GTO) dominate launch frequency
- Landing success increases with flight number for most orbits

Payload vs. Orbit Type



- Successful landings occur across multiple orbit types
- Lower payload masses show more consistent success
- Heavier payload missions are concentrated in specific orbits

Launch Success Yearly Trend



- Early years show low or zero landing success
- Success rate increases steadily after 2014

All Launch Site Names

- There are 4 distinct Launch Site

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

SQL QUERY to obtain the result

```
SELECT DISTINCT "Launch_Site" AS
Launch_Site FROM SPACEXTABLE
```

Launch Site Names Begin with 'KSC'

- 5 records where launch sites' names start with `KSC`

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
1	2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2	2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
3	2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
4	2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

SQL QUERY to obtain the result

```
SELECT * FROM SPACEXTABLE
WHERE "Launch_Site" LIKE "KSC%" LIMIT 5
```

Total Payload Mass

- The total payload carried by boosters from NASA is equal to **48213 kg**

SQL QUERY to obtain the result
SELECT SUM("Payload_Mass__kg_") AS total_payload_mass_kg FROM SPACEXTABLE WHERE "Customer" LIKE "NASA (CRS)%"

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is **2928.4kg**

SQL QUERY to obtain the result
SELECT AVG("Payload_Mass__kg_") AS avg_payload_mass_kg FROM SPACEXTABLE WHERE "Booster_Version" = "F9 v1.1"

First Successful Ground Landing Date

- The dates of the first successful landing outcome on drone ship is **2016-04-08**

SQL QUERY to obtain the result
SELECT MIN("Date") AS first_success_droneship_date FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE "Success (drone ship)%

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters Name which have successfully landed on drone ship and had payload mass between 4000 and 6000 kg

SQL QUERY to obtain the result

```
SELECT DISTINCT "Booster_Version"FROM
SPACEXTABLEWHERE "Landing_Outcome" LIKE
"Success (ground pad)%"  AND
"Payload_Mass__kg_" > 4000  AND
"Payload_Mass__kg_" < 6000
```

	Booster_Version
0	F9 FT B1032.1
1	F9 B4 B1040.1
2	F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

SQL QUERY to obtain the result

```
SELECT "Mission_Outcome",
COUNT(*) AS total
FROM SPACEXTABLE
GROUP BY "Mission_Outcome"
```

	Mission_Outcome	total
0	Failure (in flight)	1
1	Success	98
2	Success	1
3	Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

SQL QUERY to obtain the result

```
SELECT DISTINCT "Booster_Version“
FROM SPACEXTABLE
WHERE "Payload_Mass__kg_" = (SELECT
MAX("Payload_Mass__kg_") FROM
SPACEXTABLE)
```

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

2015 Launch Records

- List of succesful landing_outcomes in 2017, in ground pad including booster versions, launch_site and month

SQL QUERY to obtain the result

```
SELECT substr("Date", 6, 2) AS month,
       "Landing_Outcome",
      "Booster_Version",
       "Launch_Site“
FROM SPACEXTABLE
WHERE substr("Date", 0, 5) = "2017"  AND
"Landing_Outcome" LIKE "Success (ground
pad)%
```

	month	Landing_Outcome	Booster_Version	Launch_Site
0	02	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
1	05	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
2	06	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
3	08	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
4	09	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
5	12	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Failures are less frequent and diversified, occurring via drone-ship failures, ocean landings, or parachute issues.
- Most missions either did not attempt a landing or landed successfully

	Landing_Outcome	outcome_count
0	No attempt	10
1	Success (drone ship)	5
2	Failure (drone ship)	5
3	Success (ground pad)	3
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Failure (parachute)	2
7	Precluded (drone ship)	1

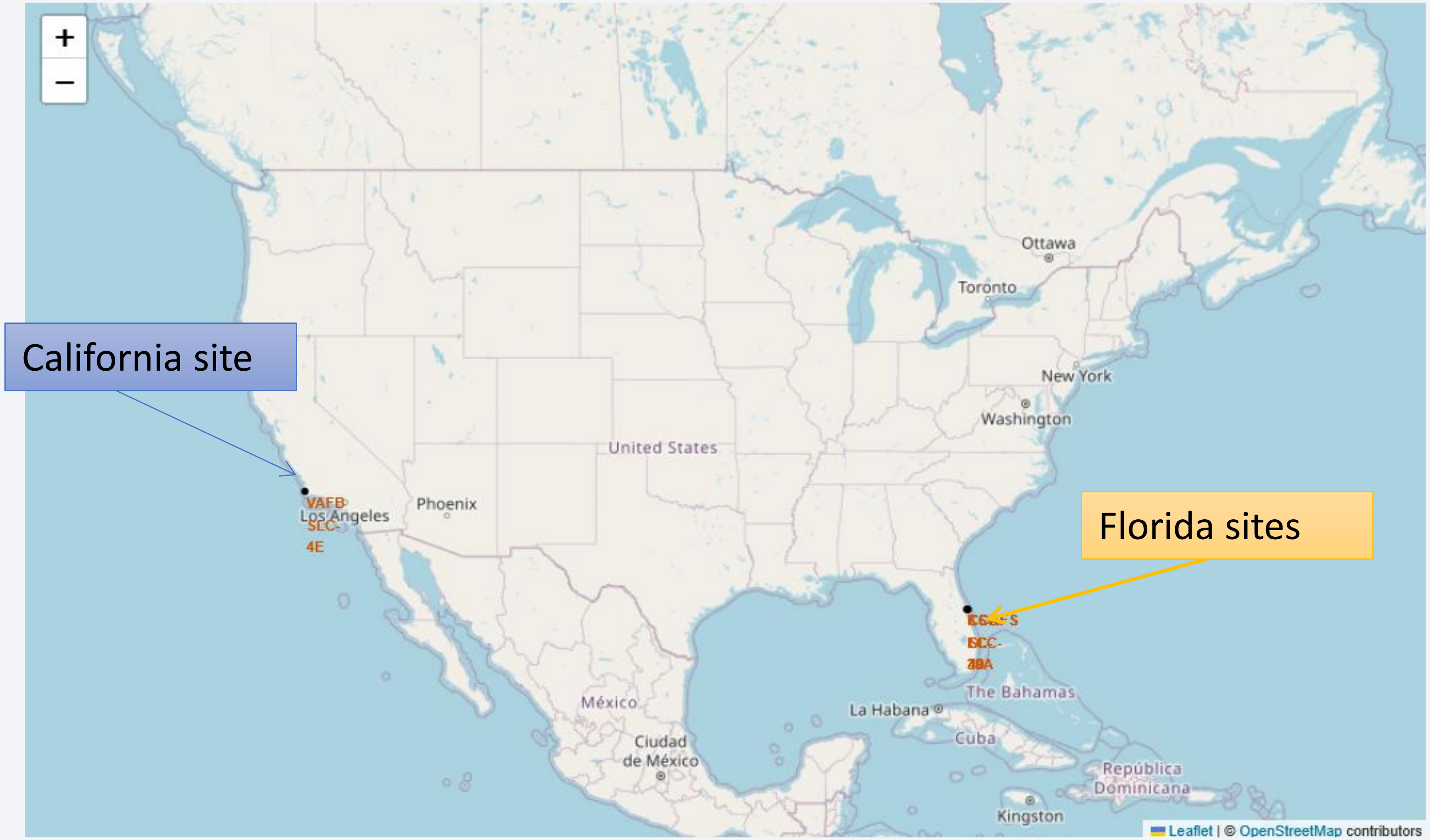
SQL QUERY to obtain the result

```
SELECT "Landing_Outcome",
      COUNT(*) AS outcome_count
FROM SPACEXTABLE
WHERE "Date" BETWEEN "2010-06-04"
AND "2017-03-20“
GROUP BY "Landing_Outcome“
ORDER BY outcome_count DESC
```

Section 3

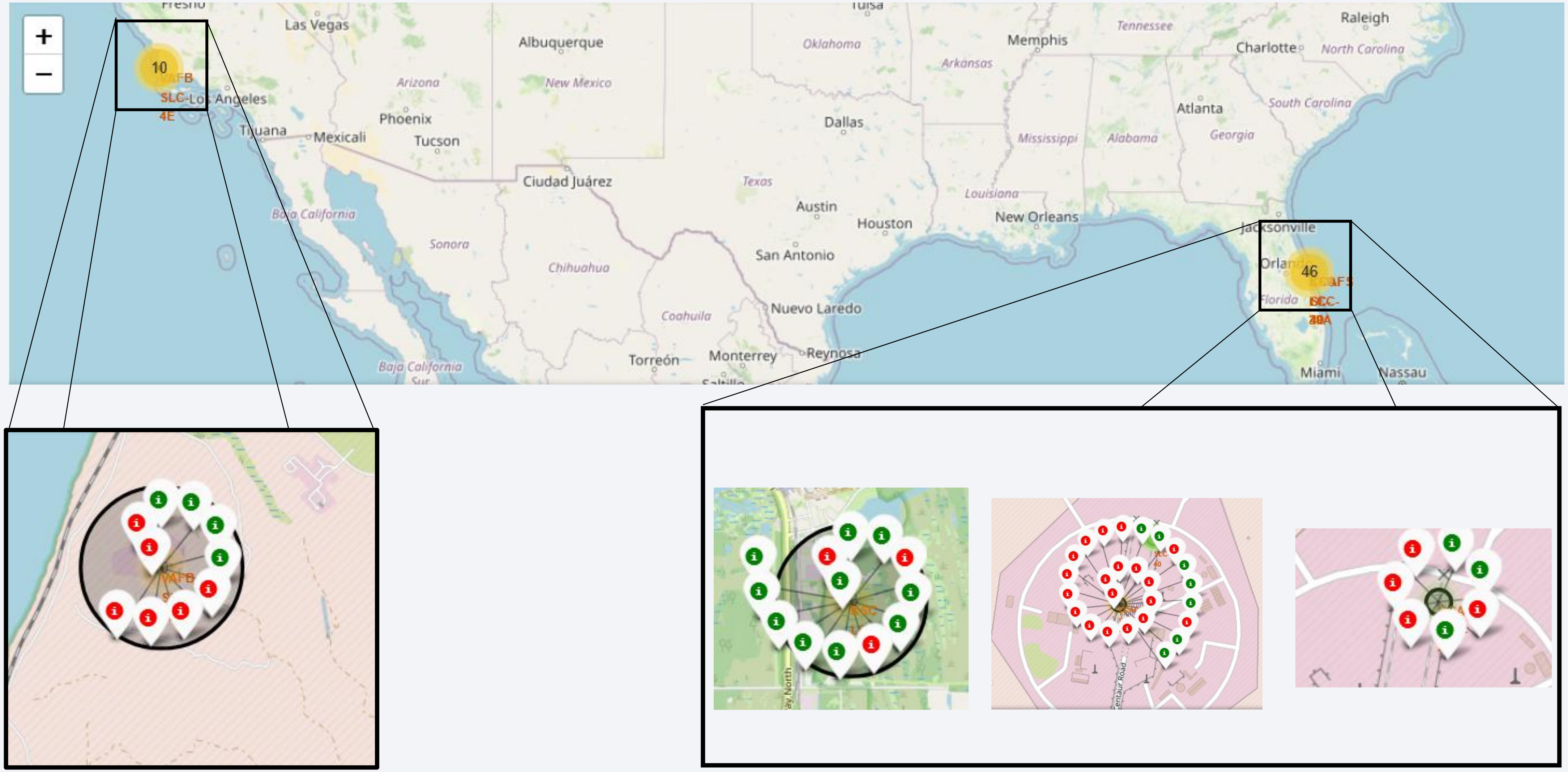
Launch Sites Proximities Analysis

SpaceX Launch Sites Location

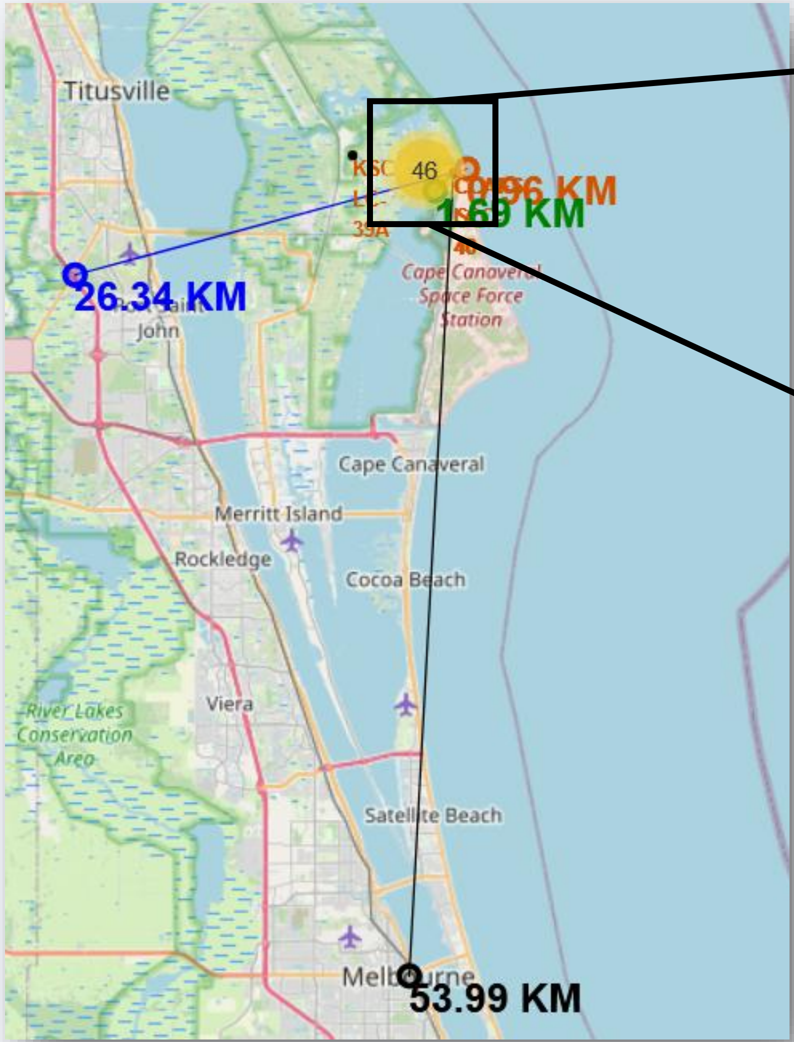


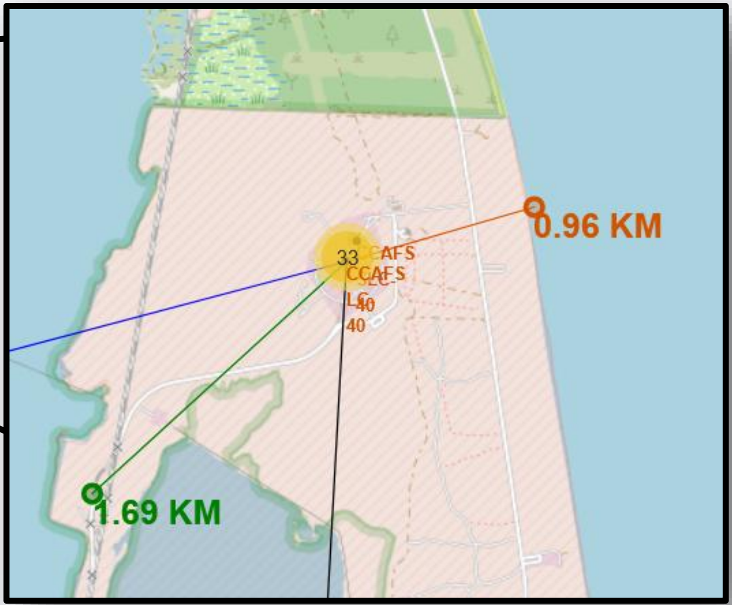
- SpaceX have 4 Launch sites on in USA
- 3 in Florida and 1 in California,
- all of them are near the coast
- Located in southern part of the country as close as possible from equator line

Launch sites with success/unsuccess label



Launch Site distance to landmarks



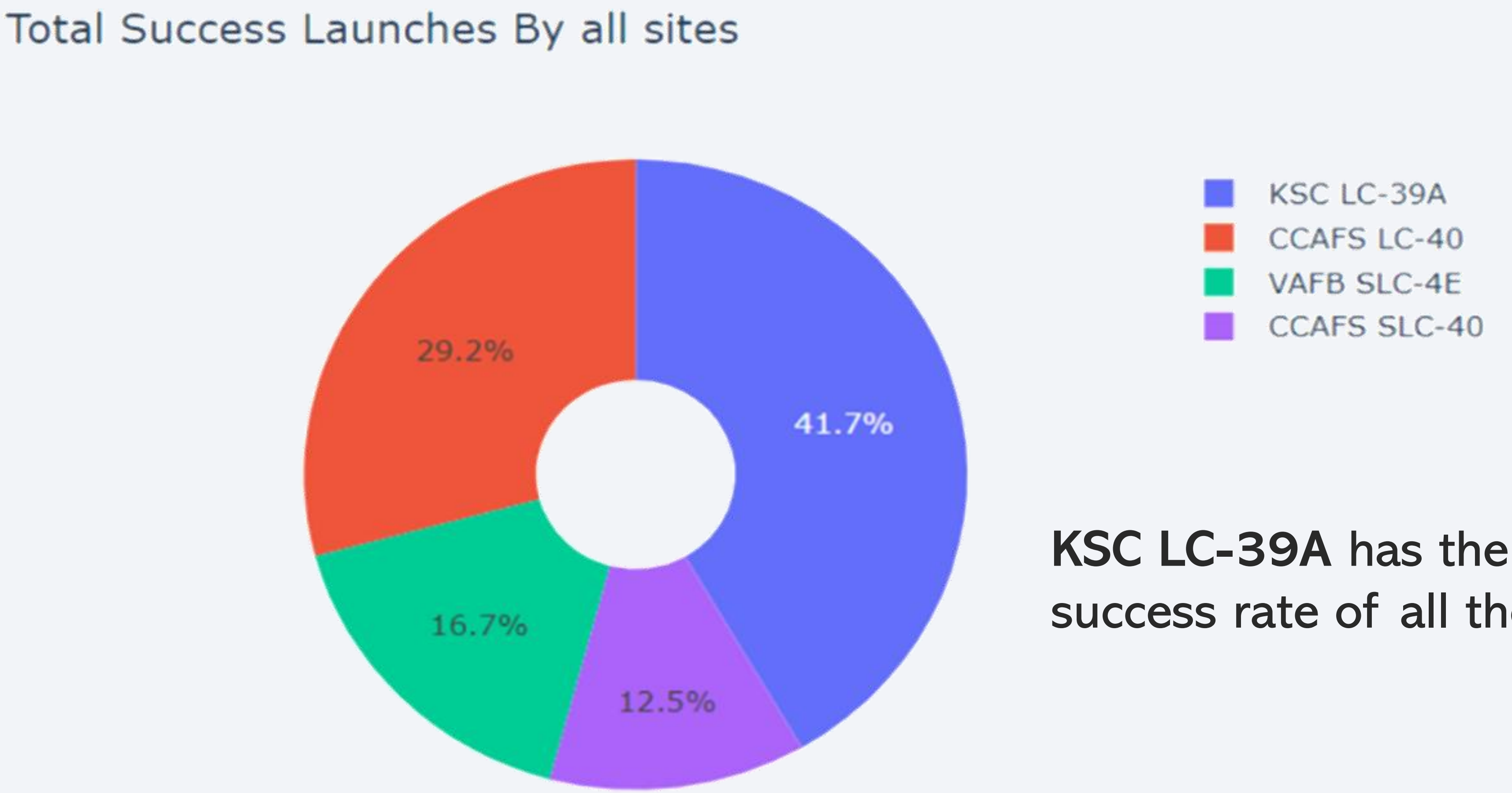


- Launch site are at close proximity to railways: 1.69km
- Launch site are at proximity to highways: 26.34Km
- Launch site are at close proximity to coastline: ~1km
- Keep distance from city, distance away from cities : 53.99km

Section 4

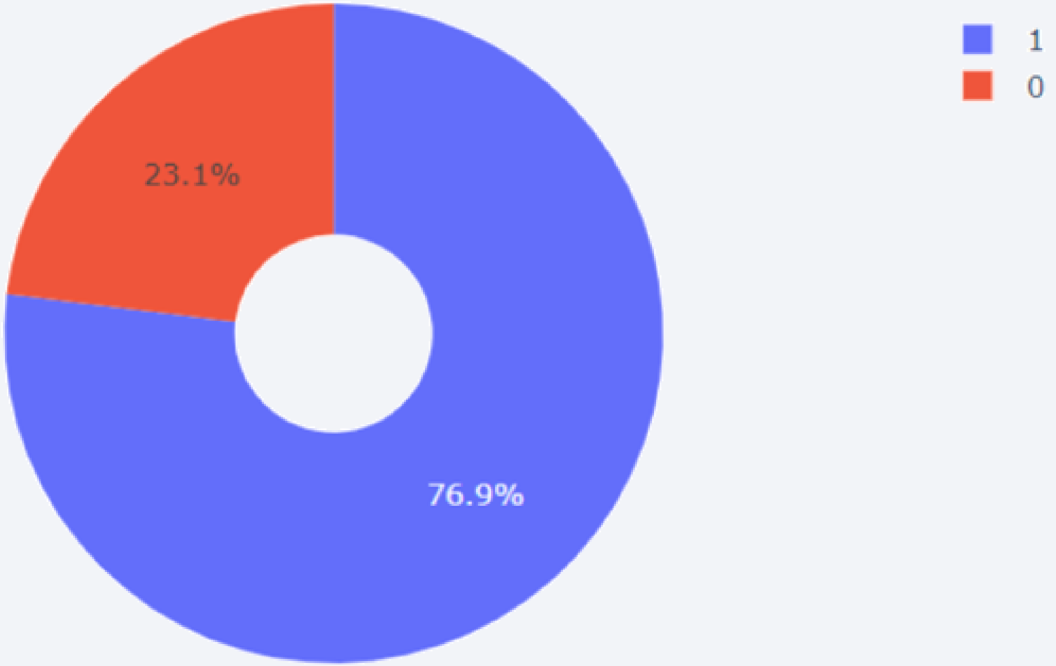
Build a Dashboard with Plotly Dash

Success rate by launch site



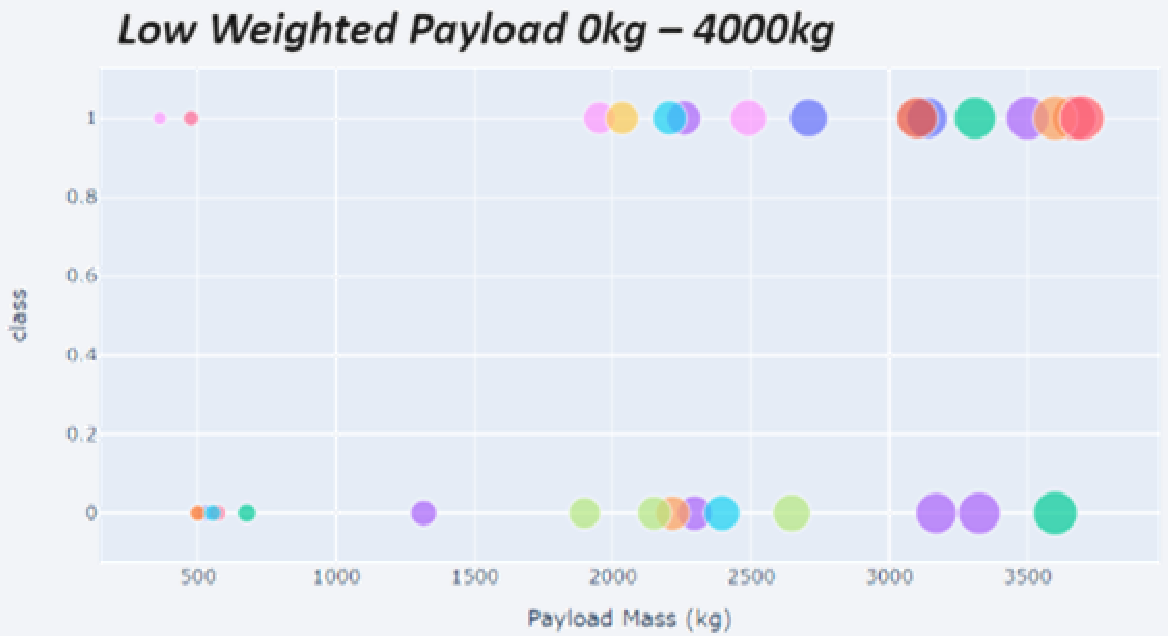
KSC LC-39A has the highest success rate of all the sites

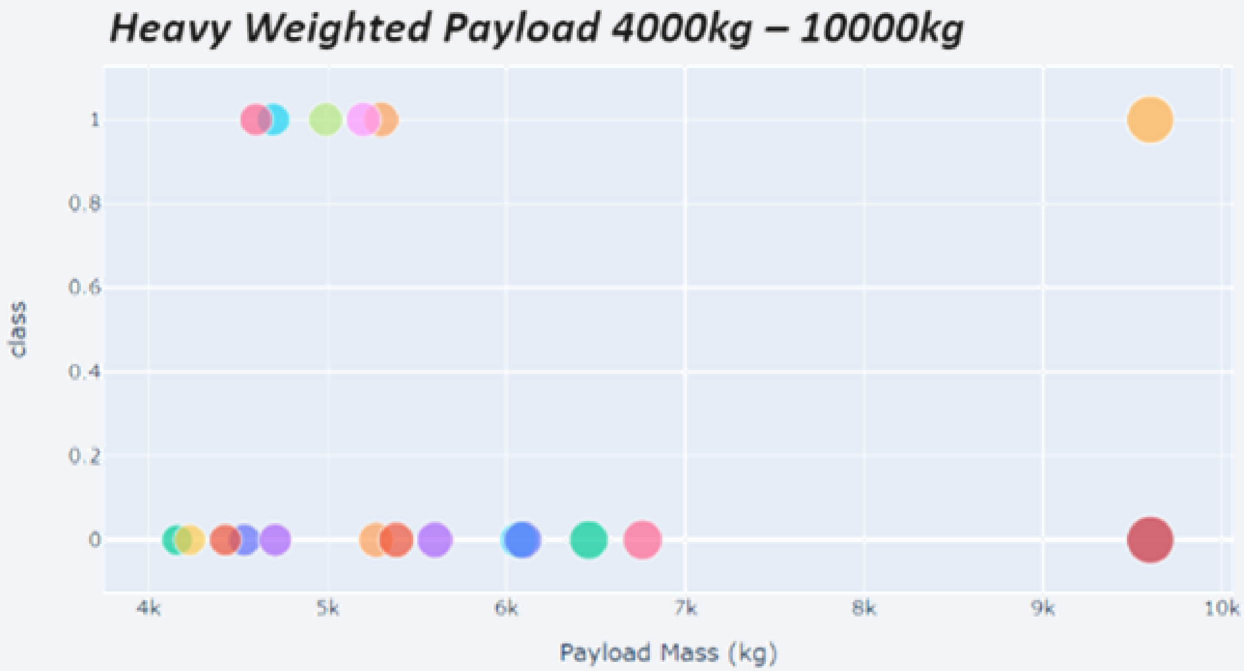
Launch site with highest success ratio



KSC LC-39A has a success ratio of 76.9%,

Payload vs. Launch Outcome



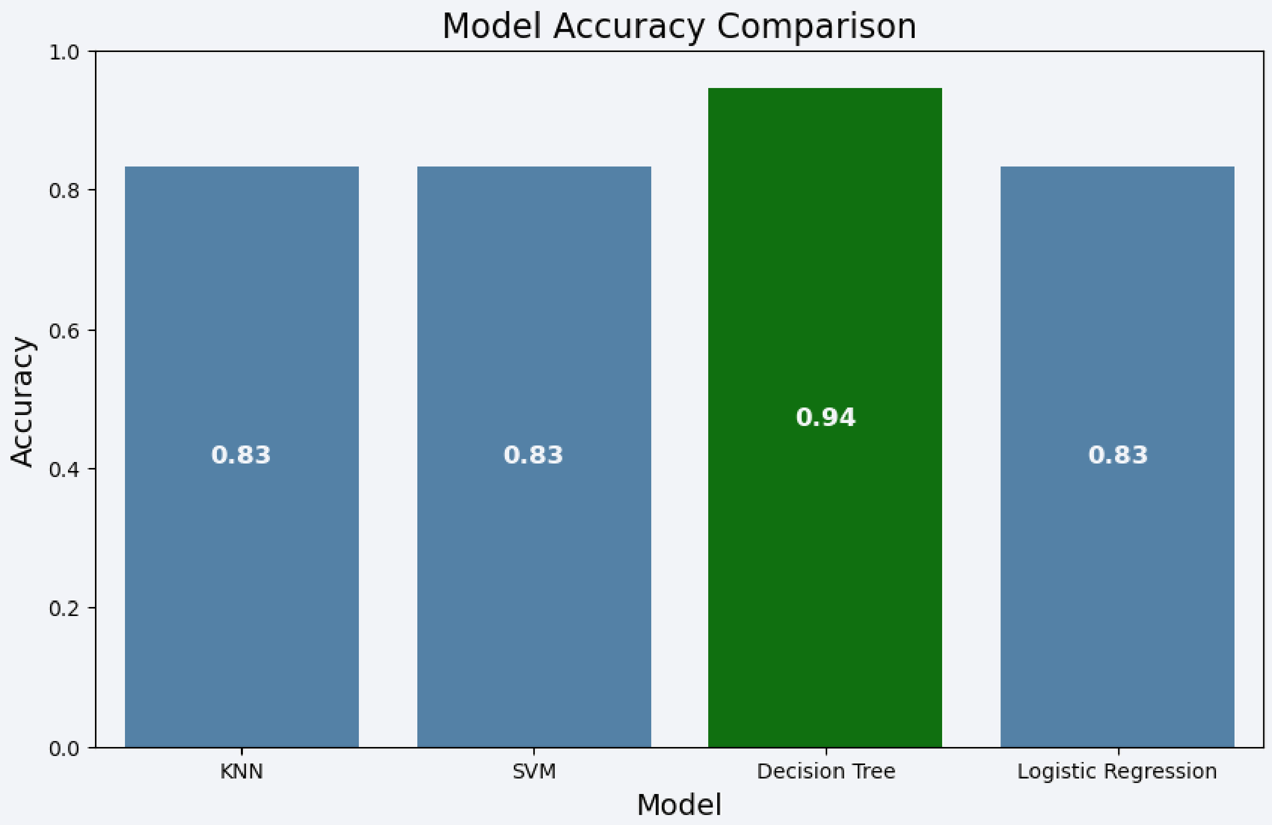


- Low-payload missions (0–4,000 kg) show a high successful landings .
- Heavy-payload missions (4,000–10,000 kg) exhibit a higher risk of landing failure

Section 5

Predictive Analysis (Classification)

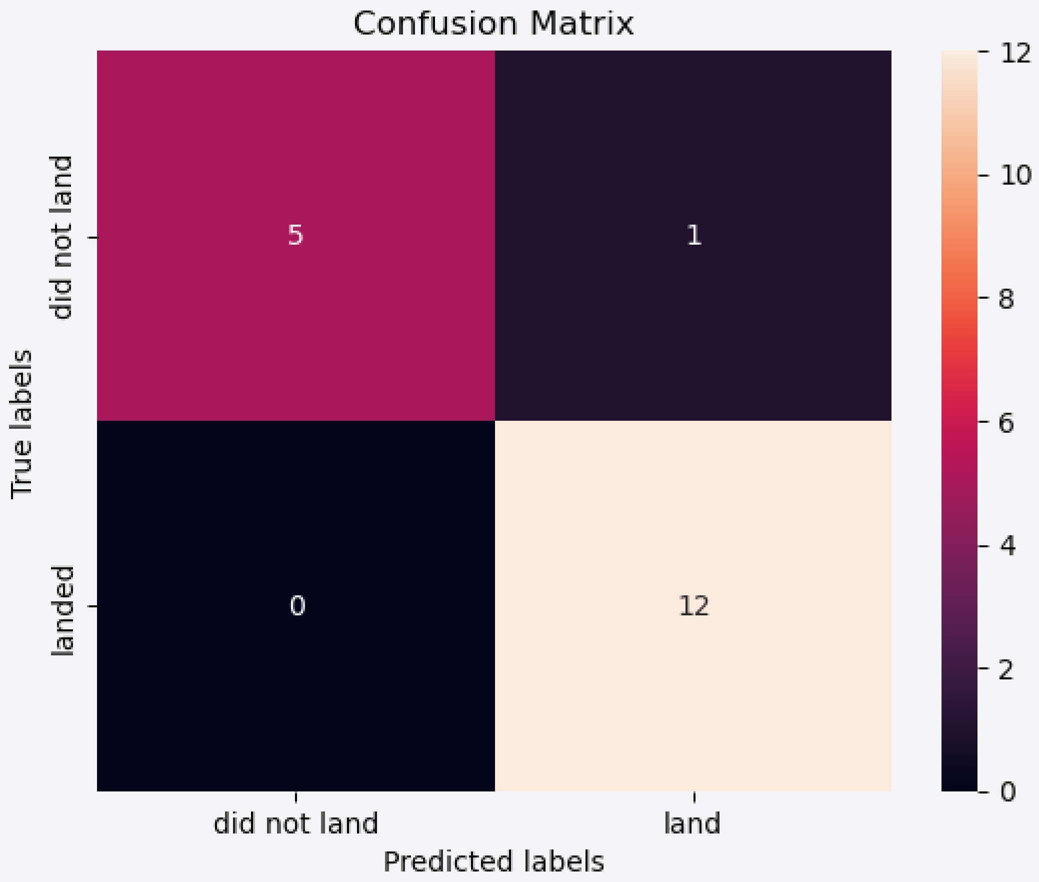
Classification Accuracy



- Using the method **score** on the test data, Decision Tree has the highest classification accuracy with a value of 0.944

Confusion Matrix

- The model is excellent at detecting successful landings (no false negatives).
- There is only one false alarm where a failure was predicted as a success.
- Overall, the classifier performs very strongly, especially for the critical class (landed).



Conclusions

- Launch success rate started to increase in over time.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

Thank you!